



Proving Positive Almost-Sure Termination

Olivier Bournez, Florent Garnier

► To cite this version:

Olivier Bournez, Florent Garnier. Proving Positive Almost-Sure Termination. [Intern report] A04-R-409 || bournez04h, 2004, 16 p. inria-00099867

HAL Id: inria-00099867

<https://hal.inria.fr/inria-00099867>

Submitted on 26 Sep 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proving Positive Almost-Sure Termination

Olivier Bournez, Florent Garnier

LORIA/INRIA, 615 Rue du Jardin Botanique
54602 Villers lès Nancy Cedex, France

Abstract In order to extend the modeling capabilities of rewriting systems, it is rather natural to consider that the firing of rules can be subject to some probabilistic laws. Considering rewrite rules subject to probabilities leads to numerous questions about the underlying notions and results.

We focus here on the problem of termination of a set of probabilistic rewrite rules. A probabilistic rewrite system is said almost surely terminating if the probability that a derivation leads to a normal form is one. Such a system is said positively almost surely terminating if furthermore the mean length of a derivation is finite. We provide several results and techniques in order to prove positive almost sure termination of a given set of probabilistic rewrite rules. All these techniques subsume classical ones for non-probabilistic systems.

1 Introduction

Since 30 years, term rewriting has shown to be a very powerful tool in any context where efficient methods for reasoning with equations are required [1,9]. In the last decade, term rewriting has also shown to provide a very elegant framework for specifying concurrency models and deduction systems [12,13].

When specifying probabilistic systems, it is rather natural to consider that the firing of a rewrite rule can be subject to some probabilistic rules. For that purpose, we proposed in [3] to add basic probabilistic strategies to rule based languages. The idea of adding probabilities to rewrite rules have also been explored in [7] in the context of probabilistic constraint handling rules, or in [14].

Considering rewrite rules subject to probabilities leads to numerous questions about the underlying notions and results. In [3], we introduced probabilistic abstract reduction systems, and we introduced notions like almost-sure termination or probabilistic confluence, with relations between all these notions. In [2], we proved that, unlike what happens for classical rewriting logic, there is no hope to build a sound and complete proof system with probabilities in the general case. We however proposed a rather natural notion of rewriting logic which is sound and complete when proof terms are explicit [2].

This paper is a contribution devoted to a next step: understand and provide proof techniques for proving termination of a set of probabilistic rewrite rules.

As in [3], we propose to call a deterministic probabilistic rewrite system almost surely terminating if the probability that a term leads to a normal form

is one. However, unlike in [3], we also allow non-deterministic systems. A non-deterministic probabilistic rewrite system is said almost surely terminating if the probability that a term leads to a normal form is one whatever the reduction strategy is.

Termination is a desirable interesting notion. However, in the probabilistic context, we think we should distinguish “reasonable” termination from general termination.

Indeed, consider a system like a symmetric random walk on the set \mathbb{Z}^k of integers. For $k = 1$ or 2 , it visits almost surely all the points [5,4]. Hence, whatever the current position is, if one wants to go to the origin, a strategy is to evolve like a symmetric random walk and stop at the origin. However, even if one is almost sure to reach the origin, the expected time before reaching the origin is infinite [4,5].

Practically speaking, that is why people wanting to go to Italy, do not usually take roads randomly but prefer rather using maps, even if a proverb says that all roads go to Rome.

Coming back to termination, the point is that in an almost surely terminating system, with probability one a term leads to a normal form, but if the mean number of a derivation is infinite then this information is rather useless.

Hence, we believe that the following notion is more interesting: a system will be said positively almost surely terminating if the mean length of a derivation is finite. After formally introducing all these notions, we will see that positive almost sure termination implies almost sure termination. The rest of this paper is then devoted to proof techniques that can be used to prove positive almost sure termination.

In particular, in the classical non-probabilistic case, a simple and often used criteria for proving termination consists in embedding the underlying abstract reduction system into the set of natural integers, in such a way that each transition corresponds to a decreasing transition. This technique is sound in the general case, and is complete for finitely branching systems [1].

We show that this technique has an equivalent for probabilistic abstract reduction systems: we prove that a probabilistic abstract reduction system is positively almost sure terminating if it can be embedded into the set of non-negative reals in such a way that each transition corresponds to a decreasing transition in mean. The technique is proved sound in the general case, and complete for finitely branching systems.

Benefiting from the possibility of considering non-deterministic probabilistic abstract reduction systems, we then define probabilistic rewrite systems. The idea is to allow in right hand sides of probabilistic rules a distribution on classical right hand sides of classical rewrite rules.

We then discuss the equivalent of the classical result that says that a rewrite system is terminating iff there is a reduction order monotone on each rewrite rule.

The paper is organized as follows: in Section 2, we recall classical non-probabilistic theory. Sections 3 and 4 recall basic probability and Markov chain

theory, and Foster's theorem respectively. Section 5 introduces probabilistic abstract reduction systems. Section 6 defines positive almost sure termination. Section 7 provides techniques for proving positive almost sure termination of a probabilistic abstract reduction system. Probabilistic rewrite systems are introduced in Section 8. Techniques for proving their positive almost sure termination are discussed in Section 9.

2 Termination and Abstract Reduction Systems

We first come back to the classical setting: see for example [1,9]. An *abstract reduction system* (ARS) is $\mathcal{A} = (A, \rightarrow)$ consisting of a set A and a binary relation $\rightarrow \subset A \times A$ on A . A *derivation* is a finite, or infinite sequence $\pi = \pi_0 \rightarrow \pi_1 \cdots \rightarrow \pi_n$ with $(\pi_i, \pi_{i+1}) \in \rightarrow$ for all i . An abstract reduction system is said *terminating* iff there is no infinite chain $a_0 \rightarrow a_1 \rightarrow \cdots$.

As said in [1], the most basic method for proving termination of some $\mathcal{A} = (A, \rightarrow)$ is to embed it into another abstract reduction system $\mathcal{B} = (B, >)$ which is known to terminate. This requires a monotone mapping $V : A \rightarrow B$, where monotone means that $x \rightarrow x'$ implies $V(x) > V(x')$. Now \rightarrow terminates because an infinite chain

$$a_0 \rightarrow a_1 \rightarrow \cdots$$

would induce an infinite chain

$$V(x_0) > V(x_1) > \cdots$$

The most popular choice for termination proofs is an embedding into $(\mathbb{N}, >)$. Its popularity comes partly from the following easy completeness result [1].

Proposition 1. *A finitely branching abstract reduction system terminates if and only if there is a monotone embedding into $(\mathbb{N}, >)$.*

As in [1], observe that the technique is sound in the general case, but complete only for finitely branching systems. Indeed, the system with $A = \mathbb{N}^2$ and \rightarrow is defined by $(i+1, j) \rightarrow (i, k)$, $(i, j+1) \rightarrow (i, j)$, for all i, j, k , is terminating, whereas there is no monotone embedding from $(\mathbb{N}^2, \rightarrow)$ to $(\mathbb{N}, >)$ [1].

3 Stochastic Sequences and Markov Chains

Let us first come back to school [8,5,16]: a σ -algebra on a set Ω is a set of subsets of Ω which contains the empty-set, and is stable by countable union and complementation. In particular, the set of subsets is a natural σ -algebra for any countable set. A *measurable space* (Ω, σ) is a set with a σ -algebra on it. A *probability* is a function P from a σ -algebra to $[0, 1]$, which is countably additive, and such that $P(\Omega) = 1$. A triplet (Ω, σ, P) is called a *probability space*.

If (Ω, σ) and (Ω', σ') are measurable spaces, a function $f : \Omega \rightarrow \Omega'$ is *measurable* if for all W in σ' , $f^{-1}(W) \in \sigma$. A *random variable* is a measurable function

on some probability space. The *mean* of a random variable V taking values in the set \mathbb{N} of integers is $E[V] = \sum_i iP(V = i)$. This value is always defined, even if it can be finite or infinite. Observe that such a random variable always satisfy the so-called *telescope formula* $E[X] = \sum_{n=0}^{\infty} P(X > n)$ [4]. For a random variable V taking values in $\mathbb{N} \cup \{+\infty\}$, the mean $E[V]$ can still always be defined: practically, it is infinite if $P(V = +\infty) > 0$ and equal to $E[V] = \sum_i iP(V = i)$ (which may still be infinite) otherwise.

Given $A, B \in \sigma$, when $P(B) > 0$, the *conditional probability of A given B* is by definition $P(A|B) = P(A \cap B)/P(B)$. The mean of random variable $V : \Omega \rightarrow \mathbb{N}$ conditioned by B is defined by $E[V|B] = \sum_i iP(V = i|B)$.

A *stochastic sequence on a set A* is a family $(X_i)_{i \in \mathbb{N}}$, of random variables defined on some fixed probability space (Ω, σ, P) with values on A . It is said to be *Markovian* if its conditional distribution function satisfies the so-called Markov property, that is for all n and $s \in A$,

$$P(X_n = s | X_0 = \pi_0, X_1 = \pi_1, \dots, X_{n-1} = \pi_{n-1}) = P(X_n = s | X_{n-1} = \pi_{n-1}),$$

and *homogeneous* if furthermore this probability is independent of n .

The matrix $(p_{s,t}^i) = (P(X_{i+1} = t | X_i = s))$ is what is called a stochastic matrix (even when A is an infinite set) [4]. It has the nice property that columns sum to 1.

Giving a *Markov Chain* is of course equivalent to giving the sequence of its stochastic matrices. Given a *Homogeneous Markov Chain* corresponds to giving a unique stochastic matrix (at any rank, the matrix is the same).

4 Foster's theorem

We are searching criteria in the spirit of Proposition 1. For that purpose, we now state the following result, that can be attributed to Foster [6]. It has strong connections with Martingale theory and can be seen as a consequence of very general results of (super) Martingale theory. However, it can be proved independently, and the proof here is adapted from [4].

Theorem 1 (Foster's Theorem).

Given a homogeneous Markov chain over a countable space A whose matrix is $P = (p_{t,s})$, if there exists a measurable subset $C \subset A$, and some function $V : A \rightarrow \mathbb{R}$, with

$$\inf_{i \in A} V(i) > -\infty$$

and such that the mean drift defined by

$$\Delta V(i) = \sum_{k \in A} p_{i,k} V(k) - V(i)$$

satisfies for some $\epsilon > 0$

$$\Delta V(i) \leq -\epsilon \quad \text{for all } i \notin C,$$

then almost surely one reaches C .

Furthermore, the mean time to reach C is finite and less than $V(i)/\epsilon$.

Proof. We use the notation X_0^n for (X_0, \dots, X_n) . Since $\inf_i V(i) > -\infty$, one may assume with loss of generality that $V \geq 0$ by adding a constant if necessary.

Call τ the return time to C , defined as $\tau_C = +\infty$ if $X_n \notin C$ for all n , and $\tau_C = \min\{n \geq 0, X_n \in C\}$ otherwise. τ_C is what is called a *stopping time*: it is a random variable that takes value in $N \cup \{+\infty\}$, and for all n , the event $\{\tau_C = n\}$ can be expressed in terms of X_0, \dots, X_n .

When F is a set, denote 1_F for its characteristic function. Consider random variable $Y_n = V(X_n)1_{n < \tau}$. The inequality in the hypothesis is just $E[V(X_{n+1})|X_n = i] \leq V(i) - \epsilon$ for all $i \notin C$.

For an $i \notin C$,

$$\begin{aligned} E_i[Y_{n+1}|X_0^n] &= E_i[Y_{n+1}1_{n < \tau}|X_0^n] + E_i[Y_{n+1}1_{n \geq \tau}|X_0^n] \\ &= E_i[Y_{n+1}1_{n < \tau}|X_0^n] \\ &\leq E_i[V(X_{n+1})1_{n < \tau}|X_0^n] \\ &= 1_{n < \tau} E_i[V(X_{n+1})|X_0^n] \\ &= 1_{n < \tau} E_i[V(X_{n+1})|X_n] \end{aligned}$$

Where $E_i[\cdot]$ stands for $E[\cdot|X_0 = i]$, the third equality comes from the fact that $1_{n < \tau}$ is a function of X_0^n , and the fourth equality from the Markov property.

We obtain

$$E_i[Y_{n+1}|X_0^n] \leq 1_{n < \tau} V(X_n) - \epsilon 1_{n < \tau}$$

since almost surely $X_n \notin C$ for $n < \tau$.

Therefore

$$E_i[Y_{n+1}|X_0^n] \leq Y_n - \epsilon 1_{n < \tau}.$$

Taking expectations

$$0 \leq E_i[Y_{n+1}] \leq E_i[Y_n] - \epsilon P_i(\tau > n),$$

where $P_i(\cdot)$ stands for $P(\cdot|X_0 = i)$.

Iterating the above equality, and observing that Y_n is non-negative, we obtain

$$0 \leq E_i[Y_0] - \epsilon \sum_{k=0}^n P_i(\tau > k).$$

But almost surely $Y_0 = V(i)$ when $X_0 = i$, and $\sum_{k=0}^{\infty} P_i(\tau > k) = E_i[\tau]$ from telescope formula.

Therefore, we have $E_i[\tau] \leq V(i)/\epsilon$.

Notice that the technique of using Foster's theorem in order to prove convergence to some set C has similarities with techniques used in self-stabilization [17,11].

5 Probabilistic Abstract Reduction Systems

We are now ready to define probabilistic abstract reduction systems (PARS). We define PARS in a slightly modified way to [3]. The main motivation is that we want to allow non-deterministic systems.

In the same way that abstract reduction systems are also called *transition systems* in other contexts, PARS can be considered as *Markov Decision Processes* [15]. The only point is that, compared to usual definitions of Markov decision processes, we explicitly allow states to be terminal, and that we do not label transitions by actions.

The idea is that a PARS is given by some set A , and a relation that relate states to distributions on their successors.

Definition 1 (PARS). *Given some denumerable set S , we note $\text{Dist}(S)$ for the set of probability distributions on S : $\mu \in \text{Dist}(S)$ is a function $S \rightarrow [0, 1]$ that satisfies $\sum_{i \in S} \mu(i) = 1$.*

A probabilistic abstract reduction system (PARS) is a pair $\mathcal{A} = (A, \rightarrow)$ consisting of a countable set A and a relation $\rightarrow \in A \times \text{Dist}(A)$.

A PARS is said deterministic if, for all a , there is at most one μ with $a \rightarrow \mu$.

A state $a \in A$ with no μ such that $a \rightarrow \mu$ is said terminal.

We now need to explain how such systems evolve: an *history* (of length $n+1$) is a finite sequence $a_0 a_1 \cdots a_n$ of elements of the state space A . It is non-terminal if a_n is. A *policy* ϕ , that can also be called a *strategy*, is a function that maps non-terminal histories to distributions in such a way that $\phi(a_0 a_1 \cdots a_n) = \mu$ is always one (of the possibly many) distribution μ with $a_n \rightarrow \mu$. An history is said *realizable*, if for all $i < n$, if μ_i denotes $\phi(a_0 a_1 \cdots a_i)$, one has $\mu_i(a_{i+1}) > 0$.

A *derivation of \mathcal{A}* is then a stochastic sequence where the non-deterministic choices are given by some policy ϕ , and the probabilistic choices are governed by the corresponding distributions.

Formally:

Definition 2 (Derivations). *A derivation π of \mathcal{A} over policy ϕ is a stochastic sequence $\pi = (\pi_i)_{i \in \mathbb{N}}$ on $A \cup \{\perp\}$ such that for all n ,*

$$P(\pi_{n+1} = \perp | \pi_n = \perp) = 1,$$

$$P(\pi_{n+1} = \perp | \pi_n = s) = 1 \text{ if } s \in A \text{ is terminal,}$$

$$P(\pi_{n+1} = \perp | \pi_n = s) = 0 \text{ if } s \in A \text{ is non-terminal,}$$

and for all $t \in A$.

$$P(\pi_{n+1} = t | \pi_n = a_n, \pi_{n-1} = a_{n-1}, \dots, \pi_0 = a_0) = \mu(t)$$

whenever $a_0 a_1 \cdots a_n$ is a realizable non-terminal history and $\mu = \phi(a_0 a_1 \cdots a_n)$.

Several observations are in order.

Remark 1. Deterministic probabilistic abstract reduction systems correspond to probabilistic abstract reduction systems considered in [3].

Remark 2. The derivations are homogeneous and Markovian when the policy ϕ is *Markovian*, i.e. when the value of $\phi(a_0 a_1 \dots a_n)$ depends only on the value of a_n . In particular, this holds for deterministic systems.

6 Termination of a Probabilistic Abstract Reduction System

If a derivation is such that $\pi_n = \perp$ for some n , then $\pi_{n'} = \perp$ almost surely for all $n' \geq n$. Such a derivation is said *terminating*. A derivation which is not terminating is said *non-terminating*. In other words, a non-terminating derivation is such that $\pi_n \in A$ ($\pi_n \neq \perp$) for all n .

Definition 3 (Almost Sure Termination). A PARS $\mathcal{A} = (A, \rightarrow)$ will be said *almost surely (a.s) terminating* iff for any policy ϕ , the probability that a derivation $\pi = (\pi_i)_{i \in \mathbb{N}}$ under policy ϕ terminates is 1: i.e. for all ϕ , $P(\exists n | \pi_n = \perp) = 1$.

This can be restated as follows: given some policy ϕ , and some state a , consider the random variable $\tau[a, \phi]$ associated to a derivation π with $\pi_0 = a$, taking values in $\mathbb{N} \cup \{+\infty\}$, defined as $+\infty$ if $\pi_n \neq \perp$ for all n , and defined as $\tau[a, \phi] = \min\{n | \pi_n = \perp\}$ otherwise. Of course, $\tau[a, \phi]$ corresponds to the number of derivations from a under strategy ϕ before termination. $\tau[a, \phi]$ is easily proved to be a stopping time for all ϕ and a .

Previous definitions can then be stated as follows:

Proposition 2. A PARS \mathcal{A} is almost surely terminating iff for all strategy ϕ and all state a , $P(\tau[a, \phi] = +\infty) = 0$.

As discussed in the introduction, this notion of termination is too weak. Even if $P(\tau[a, \phi] = \infty) = 0$, it might happen that the mean time before termination

$$T[a, \phi] = E[\tau(a, \phi)]$$

is not finite, and one may never expect to reach Italy.

That is why, we suggest to introduce the notion of positive almost sure termination. Note that the choice of the name “positive” is inspired by the distinction between positive recurrence and null recurrence in Markov chains theory [4].

Definition 4 (Positive Almost Sure Termination). A PARS $\mathcal{A} = (A, \rightarrow)$ will be said *positively almost surely (+a.s.) terminating* if for all policy ϕ , for all state $a \in A$, $T[a, \phi]$ is finite.

By the discussion in Section 3 on random variables taking values in $\mathbb{N} \cup \{+\infty\}$, we know that if $P(\tau(a, \phi) = \infty) > 0$ then necessarily $E[\tau(a, \phi)]$ is infinite. That means:

Proposition 3. *A positively almost surely terminating PARS is almost surely terminating.*

Remark 3. The previous notions subsume classical ones. As one may expect, non-probabilistic systems are special cases of probabilistic systems: an abstract reduction system is a probabilistic abstract reduction system where all the distributions are Dirac distributions. I.e. all the distributions μ have value 1 on a single point, and value 0 everywhere else. Strategies for abstract reduction systems do indeed correspond to strategies for corresponding probabilistic abstract reduction systems. Terminating derivations for abstract reduction systems do indeed correspond to terminating derivations for corresponding probabilistic abstract reduction systems. An abstract reduction system is terminating iff the corresponding probabilistic abstract reduction system is ((positively) almost surely) terminating. Note that positive almost sure termination corresponds to almost sure termination and to termination for those systems.

7 Proving positive almost sure termination

We are now going to discuss techniques for proving positive almost sure termination of a probabilistic abstract reduction system. We propose a technique that subsumes the technique of Proposition 1.

One must understand that it is not at all a coincidence, but more or less unavoidable: a deep consequence of remark 3 is that any technique for proving positive almost surely termination of probabilistic abstract reduction systems must also work for abstract reduction systems, and hence necessarily subsumes a technique for non-probabilistic abstract reduction systems.

First, we prove soundness of our technique

Theorem 2 (Soundness). *A PARS $\mathcal{A} = (A, \rightarrow)$ is +a.s. terminating if there exist some function $V : A \rightarrow \mathbb{R}$, with $\inf_{i \in A} V(i) > -\infty$, and some $\epsilon > 0$, such that, for all state $a \in A$, for all μ with $a \rightarrow \mu$, the drift in a according to μ defined by*

$$\Delta_\mu V(a) = \sum_i \mu(i)V(i) - V(a)$$

satisfies

$$\Delta_\mu V(a) \leq -\epsilon.$$

Proof. We would like to use Theorem 1. However, we can not work directly on the PARS, since even if we fix a strategy, a PARS is not necessarily an homogeneous Markov chain (the fixed policy can be non-Markovian).

The solution is to fix a policy ϕ and to work on an homogeneous Markov chain \mathcal{M}_ϕ defined on another state space: the state space of \mathcal{M}_ϕ is defined as the set of all realizable histories of \mathcal{A} .

The matrix of Markov chain \mathcal{M}_ϕ is then defined such that

- for all t , $p_{h,ht} = \mu(t)$ where $\mu = \phi(h)$ if $h = a_0a_1 \cdots a_n$ is a realizable non-terminal history, where ht stands for history $a_0a_1 \cdots a_nt$,

- $p_{h,h} = 1$ if h is a realizable terminal history,
- and every other entry of the matrix is 0.

By construction, \mathcal{M}_ϕ is an homogeneous Markov chain. Now clearly, a trajectory of PARS \mathcal{A} starting from a reaches a terminal state under policy ϕ iff the corresponding trajectory of \mathcal{M}_ϕ of same length starting from a leads to a terminal history. Furthermore, the probabilities of corresponding derivations are preserved.

Consider now function $W : S_\phi \rightarrow \mathbb{R}$ defined by

$$W(a_0 a_1 \cdots a_n) = V(a_n)$$

for all realizable history $a_0 a_1 \cdots a_n$.

We have

$$\Delta W(h) = \Delta_\mu V(h) \leq -\epsilon$$

for any non-terminal realizable history h , where $\mu = \phi(h)$.

We can then apply Theorem 1 on \mathcal{M}_ϕ , with C equal to the set of terminal realizable histories to conclude that the derivations starting from a in \mathcal{M}_ϕ reach terminal realizable histories in a time whose mean is less than $W(a)/\epsilon = V(a)/\epsilon$.

Hence, all the derivations starting from a in \mathcal{A} under policy ϕ reach terminal states in a time whose mean is also less than $V(a)/\epsilon$. This holds for all a and ϕ .

We now prove that the technique is complete for finitely branching systems.

Definition 5. *A probabilistic abstract reduction system $\mathcal{A} = (A, \rightarrow)$ is finitely branching if for all a , there is at most a finite number of distributions μ with $a \rightarrow \mu$.*

Theorem 3 (Completeness for finitely branching systems). *If a finitely branching probabilistic abstract reduction system $\mathcal{A} = (A, \rightarrow)$ is +a.s. terminating then there exist some function $V : A \rightarrow \mathbb{R}$, with $\inf_{i \in A} V(i) > -\infty$, and some $\epsilon > 0$, such that, for all state $a \in A$, for all μ with $a \rightarrow \mu$, the drift in a according to μ defined by*

$$\Delta_\mu V(a) = \sum_i \mu(i) V(i) - V(a)$$

satisfies

$$\Delta_\mu V(a) \leq -\epsilon.$$

Proof. By hypothesis, for all state a , and policy ϕ , we have $T[a, \phi] < +\infty$. When h is a realizable history, and ϕ is a policy, we write $T[h, \phi]$ for the mean time before reaching \perp after history h .

Note that for any policy ϕ , when h is a realizable non-terminal history, we have

$$T[h, \phi] = 1 + \sum_{x \in A} \phi(h)(x) T[hx, \phi] \quad (1)$$

If policy ϕ is Markovian, we have $T[hx, \phi] = T[x, \phi]$, and hence

$$T[h, \phi] = 1 + \sum_{x \in A} \phi(h)(x)T[x, \phi]. \quad (2)$$

The idea is to consider the “worst” strategy Φ . This strategy can be built as follows: in any realizable non-terminal history $h = a_0 \dots a_n$, Φ maps h to the distribution μ with $a_n \rightarrow \mu$ that maximizes $\sup_{\phi} \sum_{x \in A} \mu(x)T[hx, \phi]$.

Since to any strategy ϕ on can associate a strategy ϕ' with

$$T[hx, \phi] = T[x, \phi']$$

(take $\phi'(h') = \phi(hh')$ for any realizable non-terminal history h'),

$$\sup_{\phi} \sum_{x \in A} \mu(x)T[hx, \phi] = \sup_{\phi} \sum_{x \in A} \mu(x)T[x, \phi],$$

and hence Φ is Markovian.

We claim that this is indeed the worst strategy, i.e.

$$\sup_{\phi} T[h, \phi] \leq T[h, \Phi] \quad (3)$$

for all realizable non-terminal history h .

This follows from the following arguments: for any integer i , let Φ_i be the set of strategies that coincides with Φ on all histories of length less than i . Using repeatedly Equation 1, one gets for all integer i ,

$$\sup_{\phi} T[h, \phi] \leq \sup_{\phi \in \Phi_i} T[h, \phi]$$

for all realizable non-terminal histories h of length less than i .

Now, since $T[h, \Phi]$ is the limit of $\sup_{\phi \in \Phi_i} T[h, \phi]$ when i goes to infinity, Equation 3 holds.

Now, in any non-terminal a , with $a \rightarrow \mu$,

$$\begin{aligned} \sum_{x \in A} \mu(x)T[x, \Phi] &\leq \sup_{\phi} \sum_{x \in A} \mu(x)T[x, \phi] \\ &\leq \sup_{\phi} \sum_{x \in A} \Phi(a)(x)T[x, \phi] \\ &\leq \sum_{x \in A} \Phi(a)(x)T[x, \Phi]. \end{aligned} \quad (4)$$

where first inequality comes from the fact that Φ is a particular strategy, the second from the definition of $\Phi(a)(x)$, and the third from the fact that the sup of a sum is always less than the sum of the sups.

We are done: indeed, if we take $V(a) = T[a, \Phi]$ for all state a , and $\epsilon = 1$, we know that V is non-negative, and for any μ with $a \rightarrow \mu$, we have

$$\begin{aligned} \Delta_{\mu} V(a) &= \sum_{x \in A} \mu(x)V(x) - V(a) \\ &= \sum_{x \in A} \mu(x)T[x, \Phi] - T[a, \Phi] \\ &= -1 + (\sum_{x \in A} \mu(x)T[x, \Phi] - \sum_{x \in A} \Phi(a)(x)T[x, \Phi]) \\ &\leq -1 \end{aligned}$$

where third equality comes from Equation 2, and last inequality from Equation 4.

Remark 4. Note that the restriction to finitely branching systems in the previous theorem is mandatory: this can be seen as a consequence of Remark 3. Indeed, consider the counter-example after Proposition 1, considered as a probabilistic abstract reduction system. If there were a function V and some $\epsilon > 0$ as in the conclusion of previous theorem, adding a constant if necessary, and multiplying by $1/\epsilon$ if necessary, we can assume V non-negative, and $\epsilon = 1$. Now, in any non-terminal state a with $a \rightarrow \mu$, since we should have $\Delta_\mu V(a) \leq -1$, and since μ is a Dirac that is 0 except on some point x where it has value 1, we must have for that x , $V(x) \leq V(a) - 1$. Now, if $k = V(1, 1)$, consider the strategy going from $(1, 1)$ to $(0, k)$, $(0, k - 1)$, \dots , $(0, 0)$. V must decrease of at least 1 at each transition. That leads to a contradiction, since starting from k , one can not do it $k + 1$ times keeping V non-negative.

8 Probabilistic Rewrite Systems

We are now introducing the notion of probabilistic rewrite system. Our motivation is to get something that covers classical (i.e. non-probabilistic) rewrite systems, and also Markov chains over finite spaces. Doing so, we can claim that all examples that have been modeled in literature using finite Markov chains (for e.g. in model-checking contexts [10]) can be modeled in this framework.

Definition 6 (Probabilistic Rewrite system). *A probabilistic rewrite rule is an element of $T(\Sigma, X) \times \text{Dist}(T(\Sigma, X))$. A probabilistic rewrite system is a finite set \mathcal{R} of probabilistic rewrite rules.*

To a probabilistic rewrite system is associated an probabilistic abstract reduction $(T(\Sigma, X), \rightarrow_{\mathcal{R}})$ system over the set of terms $T(\Sigma, X)$ where $\rightarrow_{\mathcal{R}}$ is defined as follows: When $t \in T(\Sigma, X)$ is a term, let $\text{Pos}(t)$ be the set of its positions. For $\rho \in \text{Pos}(t)$, let $t|_{\rho}$ be the subterm of t at position ρ , and let $t[s]_{\rho}$ denote the replacement of the subterm at position ρ in t by s . The set of all substitutions is denoted by Sub .

Definition 7 (Reduction relation). *To a probabilistic rewrite system \mathcal{R} is associated the following PARS $(T(\Sigma, X), \rightarrow)$ over terms:*

$$t \rightarrow_{\mathcal{R}} \mu$$

iff there is a rule $(g, M) \in \mathcal{R}$, some position $p \in \text{Pos}(t)$, some substitution $\sigma \in \text{Sub}$, such that $t|_p = \sigma(g)$, and, for all t' ,

$$\mu(t') = \sum_{t' = t[\sigma(d)]_p} M(d).$$

For example, a probabilistic rewrite rule can be $f(x, y) \mapsto g(a) : 1/2 | y : 1/2$, where $g(a) : 1/2 | y : 1/2$ denotes the distribution with value $1/2$ on $g(a)$ and value $1/2$ on y . Then $f(b, c)$ rewrites to $g(a)$ with probability $1/2$, and to c with probability $1/2$. Now, $f(b, g(a))$ rewrites to $g(a)$ with probability 1.

9 Termination of a Probabilistic Rewrite System

We now provide an equivalent of the result that says that a rewrite system is terminating iff there is a reduction order monotone on each rewrite rule [1,9].

Theorem 4. *A probabilistic rewrite system \mathcal{R} is positively almost surely terminating if and only if there exists some function $V : T(\Sigma, X) \rightarrow \mathbb{R}$, with $\inf_{i \in A} V(i) > -\infty$, and some $\epsilon > 0$, such that*

1. *“the drift of each rule is less than $-\epsilon$ ”: for each probabilistic rewrite rule $g \rightarrow M \in \mathcal{R}$, the drift*

$$\Delta_M V(g) = \sum_d M(d)(V(d) - V(g))$$

satisfies

$$\Delta_M V(g) \leq -\epsilon.$$

2. *“drift being less than $-\epsilon$ is preserved by substitutions”: for each term $s \in T(\Sigma, X)$, for all μ with $s \rightarrow \mu$, for all substitution $\sigma \in \text{Sub}$, if*

$$\Delta_\mu V(s) \leq -\epsilon$$

, then the drift

$$\Delta_{\sigma(\mu)} V(\sigma(s)) = \sum_{s'} \mu(s')(V(\sigma(s')) - V(\sigma(s)))$$

satisfies

$$\Delta_{\sigma(\mu)} V(\sigma(s)) \leq -\epsilon$$

3. *“drift being less than $-\epsilon$ is preserved by contexts”: for each term $s_1, \dots, s_n, s \in T(\Sigma, X)$, for all μ with $s \rightarrow \mu$, for all function symbol f , if*

$$\Delta_\mu V(s) \leq -\epsilon,$$

then the drift

$$\Delta_{f(s_1, \dots, \mu, \dots, s_n)} V(f(s_1, \dots, s, \dots, s_n)) = \sum_{s'} \mu(s')(V(f(s_1, \dots, s', \dots, s_n)) - V(f(s_1, \dots, s, \dots, s_n)))$$

satisfies

$$\Delta_{f(s_1, \dots, \mu, \dots, s_n)} V(f(s_1, \dots, s, \dots, s_n)) \leq -\epsilon.$$

Proof. If \mathcal{R} is positively almost surely terminating, then by Theorem 3, there exists some function $V : T(\Sigma, X) \rightarrow \mathbb{R}$, with $\inf_{i \in A} V(i) > -\infty$, and some $\epsilon > 0$, such that, for all state $a \in T(\Sigma, X)$, for all μ with $a \rightarrow \mu$, $\Delta_\mu V(a) \leq -\epsilon$.

In particular, for $a = g$, we have $a \rightarrow \mu$, where $\mu(t') = M(t')$, and hence

$$\begin{aligned} \Delta_\mu V(a) &= \sum_{t'} \mu(t') V(t') - V(a) \\ &= \sum_d M(d)(V(\sigma(d)) - V(a)) \\ &= \Delta_M V(g) \\ &\leq -\epsilon. \end{aligned}$$

Now, when $s \rightarrow \mu'$, for $a = \sigma(s)$, we have $a \rightarrow \mu$, where $\mu(\sigma(s')) = \mu'(s')$, and hence

$$\begin{aligned}\Delta_\mu V(a) &= \sum_{t'} \mu(t')V(t') - V(a) \\ &= \sum_{t'=\sigma(s')} \mu'(s')V(\sigma(s')) - V(a) \\ &= \Delta_{\sigma(\mu')} V(\sigma(s)) \\ &\leq -\epsilon.\end{aligned}$$

In a same way, when $s \rightarrow \mu'$, for $a = f(s_1, \dots, s, \dots, s_n)$, we have $a \rightarrow \mu$ where $\mu(f(s_1, \dots, s', \dots, s_n)) = \mu'(s')$, and hence

$$\begin{aligned}\Delta_\mu V(a) &= \sum_{t'} \mu(t')V(t') - V(a) \\ &= \sum_{s'} \mu'(s')V(f(s_1, \dots, s', \dots, s_n)) - V(f(s_1, \dots, s, \dots, s_n)) \\ &= \Delta_{f(s_1, \dots, \mu', \dots, s_n)} V(f(s_1, \dots, s, \dots, s_n)) \\ &\leq -\epsilon.\end{aligned}$$

This proves that conditions 1, 2 and 3 are necessary.

Conversely, assume that conditions 1, 2 and 3, hold. We have $t \rightarrow \mu$ iff there is a rule $(g, M) \in \mathcal{R}$, some position $p \in \text{Pos}(t)$, some substitution $\sigma \in \text{Sub}$, such that $t|_p = \sigma(g)$, and, for all t' , $\mu(t') = \sum_{t'=t[\sigma(d)]_p} M(d)$.

Since a derivation $t \rightarrow \mu$ is necessarily via some rule (g, M) , from Theorem 2, we only need to prove that for all rule (g, M) and term t , if $t \rightarrow \mu$ via (g, M) then $\Delta_\mu V(t) \leq -\epsilon$.

This is proved by induction on the length of p . If p is of length 0, then $t = \sigma(g)$. By condition 1, we know that $\Delta_M V(g) \leq -\epsilon$. By condition 2, since $g \rightarrow M$, and $\Delta_M V(g) \leq -\epsilon$, we have $\Delta_\mu V(t) = \Delta_{\sigma(M)} V(\sigma(g)) \leq -\epsilon$, where the equality is established as in the third paragraph above.

If $p = p_1 p_2 \dots p_k$ is of length $k > 0$, then t can be written as $f(s_1, \dots, s, \dots, s_n)$ and $s \rightarrow \mu'$ via (g, M) . By induction hypothesis, $\Delta_{\mu'} V(s) \leq -\epsilon$. By condition 3, $\Delta_\mu V(t) = \Delta_{f(s_1, \dots, \mu', \dots, s_n)} V(f(s_1, \dots, s, \dots, s_n)) \leq -\epsilon$, where the equality is established as in the fourth paragraph above.

Sufficient conditions for 1, 2 and 3 can be established. Indeed:

Definition 8 (Substitution and Context decrease). *Given a set $S \subset T(\Sigma, X)$ of terms.*

- A function $V : T(\Sigma, X) \rightarrow \mathbb{R}$ is *substitution decreasing on S* if for all $t, t' \in S$, and all substitutions $\sigma \in \text{Sub}$,

$$V(\sigma(t)) - V(\sigma(t')) \leq V(t) - V(t').$$

- A function $V : T(\Sigma, X) \rightarrow \mathbb{R}$ is *context decreasing on S* if for all $t, t' \in S$, and all s_1, \dots, s_n and function symbol f ,

$$V(f(s_1, \dots, t, \dots, s_n)) - V(f(s_1, \dots, t', \dots, s_n)) \leq V(t) - V(t').$$

Definition 9 (Terms appearing in a rule). *Given a probabilistic rewrite rule (g, M) , the set of terms appearing in the rule is the union of g and of the d with $M(d) > 0$.*

Example 1. The terms appearing in $f(x) \rightarrow f(f(x)) : p_1 | g(f(x)) : p_2 | g(x) : p_3$ are $f(x)$, $f(f(x))$, $g(f(x))$, and $g(x)$.

Theorem 5. *A probabilistic rewrite system \mathcal{R} is positively almost surely terminating if there exists some function $V : T(\Sigma, X) \rightarrow \mathbb{R}$, with $\inf_{i \in A} V(i) > -\infty$, and some $\epsilon > 0$, such that the drift of each rule is less or equal to $-\epsilon$, and V is substitution decreasing on terms appearing in each rules, and context decreasing on $T(\Sigma, X)$.*

Proof. Condition 1 holds by hypothesis.

Since V is context decreasing, for all f, s, s_1, \dots, s_n and μ , we have

$$\Delta_{f(s_1, \dots, \mu, \dots, s_n)} V(f(s_1, \dots, s, \dots, s_n)) \leq \Delta_\mu V(s)$$

and so, condition 3 holds.

Now, given conditions 1 and 3, the proof of indirect sense of Theorem 4, only require that $\Delta_{\sigma(M)} V(\sigma(g)) = \sum_d M(d)(V(\sigma(d)) - V(\sigma(g))) \leq -\epsilon$ for each probabilistic rule (g, M) and substitution σ . Since the g and d with $M(d) > 0$ are appearing in the rule (g, M) , and since V is substitution decreasing on such terms, this is less or equal to $\sum_d M(d)(V(d) - V(g))$, i.e. less or equal to $-\epsilon$ by condition 1.

We are now discussing some examples:

Example 2. The probabilistic rewrite system restricted to the unique rule

$$a \rightarrow a : 1/2 | b : 1/2$$

is +a.s. terminating. Indeed, consider $V(a) = 10$, $V(b) = 2$, and observe that $1/2 \times 10 + 1/2 \times 2 - 10 < 0$.

Example 3. The probabilistic rewrite system

$$\begin{aligned} f(x) &\rightarrow x : p_1 | f(f(x)) : 1 - p_1 \\ f(x) &\rightarrow x : p_2 | f(f(x)) : 1 - p_2 \end{aligned}$$

is +a.s. terminating if $p_1 > 1/2$ and $p_2 > 1/2$. Indeed, consider V that counts the depth in symbol f of a term. V can be shown context decreasing on all terms. V is substitution decreasing on $\{f(x), x, f(f(x))\}$. Now the drift of each rule is given by $-1 \times p_i + 1 \times (1 - p_i) = 1 - 2p_i \leq \min(1 - 2p_1, 1 - 2p_2) < 0$.

Example 4. The probabilistic rewrite system

$$\begin{aligned} f(x) &\rightarrow f(f(x)) : p_{1_1} | g(f(x)) : p_{1_2} | g(x) : p_{1_3} \\ f(f(x)) &\rightarrow h(g(f(f(x))), f(x)) : p_{2_1} | g(f(x)) : p_{2_2} | f(g(f(f(x)))) : p_{2_3} \end{aligned}$$

It is +a.s. terminating if $p_{1_1} < p_{1_3}$ and $p_{2_3} < p_{2_2}$. Indeed, consider same function V , and so context decreasing on all terms, which is substitution decreasing on $\{f(x), f(f(x)), g(f(x)), g(x)\}$ and on

$$\{f(f(x), h(g(f(f(x))), f(x))), g(f(x)), f(g(f(f(x))))\}$$

and the drifts of the rules are $1 \times p_{1_1} + 0 \times p_{1_2} - 1 \times p_{1_3} < 0$ and $0 \times p_{2_1} - 1 \times p_{2_2} + 1 \times p_{2_3} < 0$.

10 Acknowledgment

We would like to thanks Thierry Heuillard for pointing out the fact that for protocols, and in almost all other contexts, only positive almost sure termination is interesting. His remarks and suggestions motivated a great part of this work.

References

1. Franz Baader and Tobias Nipkow. *Term Rewriting and all That*. Cambridge University Press, 1998.
2. Olivier Bournez and Mathieu Hoyrup. Rewriting logic and probabilities. In Robert Nieuwenhuis, editor, *Rewriting Techniques and Applications, 14th International Conference, RTA 2003, Valencia, Spain, June 9-11, 2003, Proceedings*, volume 2706 of *Lecture Notes in Computer Science*, pages 61–75. Springer, June 2003.
3. Olivier Bournez and Claude Kirchner. Probabilistic rewrite strategies: Applications to ELAN. In Sophie Tison, editor, *Rewriting Techniques and Applications*, volume 2378 of *Lecture Notes in Computer Science*, pages 252–266. Springer-Verlag, July 22-24 2002.
4. Pierre Brémaud. *Markov Chains*. Springer, 1991.
5. W. Feller. *An introduction to probability theory and its applications, volume 1*. Wiley, 1968.
6. F. G. Foster. On the stochastic matrices associated with certain queuing processes. *The Annals of Mathematical Statistics*, 24:355–360, 1953.
7. Thom Frühwirth, Alexandra Di Pierro, and Herbert Wiklicky. Toward probabilistic constraint handling rules. In Slim Abdennadher and Thom Frühwirth, editors, *Proceedings of the third Workshop on Rule-Based Constraint Reasoning and Programming (RCoRP'01)*, Paphos, Cyprus, December 2001. Under the hospice of the International Conferences in Constraint Programming and Logic Programming.
8. G. Grimmett. *Probability Theory*. Cambridge University Press, 1993.
9. Jan Willem Klop. Term rewriting systems. In S. Abramsky, D. M. Gabbay, and T. S. E. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 2, chapter 1, pages 1–117. Oxford University Press, Oxford, 1992.
10. Marta Kwiatkowska, Gethin Norman, and David Parker. PRISM: Probabilistic symbolic model checker. *Lecture Notes in Computer Science*, 2324:200, 2002.
11. N. Lynch. *Distributed Algorithms*. Morgan Kaufmann Publishers, Inc, 1997.
12. Narciso Martí-Oliet and José Meseguer. Rewriting logic: Roadmap and bibliography. *Theoretical Computer Science*, 285(2):121–154, 2002.
13. J. Meseguer. Conditional rewriting logic as a unified model of concurrency. *Theoretical Computer Science*, 96(1):73–155, 1992.
14. Kumar Nirman, Koushik Sen, Jose Meseguer, and Gul Agha. A rewriting based model for probabilistic distributed object systems. In *Proceedings of 6th IFIP International Conference on Formal Methods for Open Object-based Distributed Systems (FMODS'03)*, volume 2884 of *Lecture Notes in Computer Science*, pages 32–46. Springer, Paris, France, November 2003.
15. M.L. Puternam. *Markov Decision Processes - Discrete Stochastic Dynamic Programming*. Wiley series in probability and mathematical statistics. John Wiley & Sons, 1994.
16. W. Rudin. *Real and Complex Analysis, 3rd edition*. McGraw Hills, USA, March 1987.
17. M. Schneider. Self-stabilization. *ACM Computing Surveys*, 25:45–67, 1993.

A Conclusion and perspective

In this paper we presented non-deterministic probabilistic abstract reduction systems, probabilistic rewrite systems, and we gave necessary and sufficient conditions for proving positive and almost sure termination of these systems. We also provided tractable sufficient conditions and application examples.

We believe that our notion of probabilistic rewrite system is very powerful since it covers all systems that can be encoded by classical rewrite systems and finite Markov chains. In particular, we already explored it to model a telecommunication protocol.

Next step include understanding whether there could be valid and interesting results generalizing techniques based on polynomial orders, or even on semantical methods.